Recent advances in understanding the replication initiator protein of the ssDNA plant viruses of the family *Nanoviridae*

Sangita Venkataraman¹ · R. Selvarajan²

Received: 12 September 2018 / Accepted: 13 February 2019 / Published online: 1 March 2019
© Indian Virological Society 2019

Abstract The families of viruses possessing single-stranded (ss) circular genome employ a dedicated replication initiator protein (Rep) for making copies of their genome through the process of rolling circle replication. The replication begins at conserved nonanucleotide sequence at the intergenic region. The Rep protein seems to be the most conserved amongst the available proteins of the nanovirids and comprises of the N-terminal endonuclease domain and the C-terminal helicase domain. The structural studies of Faba bean necrotic yellows virus endonuclease domain suggests a $\alpha + \beta$ fold comprising of central β sheet built from five antiparallel β strands surrounded by outer short α helices. The catalysis is mediated by a conserved Tyr residue and employs divalent metal ions (Mn^{2+}). On one hand, the Reps associate with each other and oligomerize and on the other hand interact with varied host and vector associated proteins for successful infection. The sequence analysis of Reps from previously known nanovirids and the newly found ones from metagenomics data shed light on the evolutionary pattern of nanovirids in comparison to other plant infecting ssDNA viruses.

Keywords Nanoviridae · Replication initiator protein · Rolling circle replication · Molecular phylogeny · Structural motifs · Interactions

Introduction

Circular ss DNA viruses of eukaryotes are amongst the smallest viruses infecting various kingdoms of life. They are currently grouped into eight families: *Bacilladnaviridae* infecting Diatoms [24], *Anelloviridae* [2] and *Sma-coviridae* [45] infecting animals and human, *Genomoviridae* infecting both plants and animals [52], *Parvoviridae* infecting vertebrate and arthropod hosts [11], *Circoviridae* infecting birds and mammals [3], and *Geminiviridae* [54] and *Nanoviridae* [31, 51] infecting plants. Except for *Anelloviridae*, the rest of the families comprise of members with multipartite genome. The important proteins encoded by these viruses are the Rep and the capsid protein (CP). The Rep is responsible for the replication of genome, while the CP forms isometric particles and aids in packing the genome into them [16, 41, 42]. The Reps are evolutionarily better conserved than the CPs and hence serve as useful markers for identification and classification of ssDNA viruses. The Reps of geminiviruses and circoviruses are closely related in sequence, structure and function to those of *Nanoviridae* [25].

The family *Nanoviridae* encompasses members that have a multipartite genome with six to eight circular segments (DNA-R, -S, -M, -C, -N and -U) ranging in size from 923 to 1111 nucleotides (nt) that encode at least 5–7 proteins and are encapsidated separately inside individual icosahedral shells of T = 1 symmetry [5, 49, 51]. The family comprises of two genera, the *Nanovirus* with type member *Subterranean clover stunt virus* (SCSV) and the

Electronic supplementary material The online version of this article (doi:[10.1007/s13337-019-00514-9](https://doi.org/10.1007/s13337-019-00514-9)) contains supplementary material, which is available to authorized users.

✉ Sangita Venkataraman
dr.sangita.b@gmail.com

✉ R. Selvarajan
selvarajanr@gmail.com

¹ Department of Biotechnology, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur 522510, India

² ICAR National Research Centre for Banana, Thayanur Post, Tiruchirapalli 620102, India

Babuvirus with type member *Banana bunchy top virus* (BBTV) [31]. While the former includes the legume infecting nanoviruses such as FBNYV, Milk vetch dwarf virus (MDV) and Pea necrotic yellow dwarf virus (PNYDV) [28] in addition to SCSV, the genus *Babuvirus* includes Abaca bunchy top virus (ABTV) and Cardamom bushy dwarf virus (CBDV) alongside BBTV [36]. The non-enveloped shell is a small round capsid of about 18–19 nm diameter and is made of sixty copies of viral CPs encoded by DNA-S. BBTV is known to infect all members of Musaceae family and is transmitted by aphids. While DNA-C, -N and -M code for the cell-cycle link (clink) protein, nuclear shuttle protein (NSP), and the movement protein, respectively, the function of DNA-U remains unknown for BBTV [53]. While the rep encoded by DNA-R, also referred to as the master-rep is known to be solely necessary and sufficient for replicating all the genomic components while the reps encoded by other genomic segments can replicate no other genomic components except its own [43]. In this review, recent developments on the Reps of nanovirids including their role in RCR, structural features, sequence analysis, and other aspects are discussed in detail.

The Rep gene

The Rep gene is about 1000-nt long and transcribed from the DNA-R segment and encodes the 33 kDa replicase protein. The GO function associated with Rep gene includes ATPase, endodeoxyribonuclease, nucleotidyl-transferase and helicase activity [13]. Though all the *Nanoviridae* DNAs contain a major sense ORF transcribed unidirectionally, the BBTV DNA-R encodes two proteins viz. the Rep and the U5 protein [51]. The Rep gene has a promoter sequence with a TATA box and an AATAAA-like polyadenylation signal at the 3' of the ORF. The position of nanoviral polyadenylation signal is between the TATA box and the ORF leading to the synthesis of a terminally redundant mRNA [17]. This mRNAs folds into extended secondary structures which are believed to aid in regulating the expression of the Rep protein. The nanoviral genome share two homologous regions namely the common region stem-loop (CR-I) and the common region II (CR-II) (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/ssdna-viruses-2011/w/ssdna_viruses/149/nanoviridae). The CR-I is conserved across individual Nanovirid genomes and contains both the origin of the virion replication and the iterated direct and inverted repeats that are recognized by the Rep during both the initiation and the resolution of RCR. The loop-forming sequence containing the nonanucleotide TATTATTAC or TAGTATTAC is highly conserved in all the species of the genus *Babuvirus*

and Nanovirus, respectively. Many monopartite and multipartite viruses have conserved origins of replication among genome segments and any mutation in this region is known to drastically affect the replication efficiency [38]. Conservation of replication origin promotes the chances of homologous-recombination between heterologous segments of different viruses thereby aiding in evolution. Sicard et al. [38] have hypothesized differential regulation of gene copy number (GCN) in multipartite viruses. It is not well understood as to how these multipartite viruses having interdependent components regulate the GCN as they are physically separated information units coding for only one component. In case of FBNYV, the relative GCN of each genomic component has been reported and it has been shown that the copy number relative ratio varied with the hosts it infects. In case of FBNYV genome formula in *Vicia faba* is 3C 3M 13N 2R 1S 7U1 10U2 16U4 and in *Medicago truncatula* the GCN is 12C 1M 1N 2R 2S 5U1 6U2 7U4. Mware [34] reported genome copy number of each genomic components of BBTV in four different banana varieties. The abundant genome component was one coding the NSP and the least abundant genome component was DNA-S. The BBTV genome formula for Dwarf Cavendish banana was 1S 2C 12M 98U3 108R 1663N while in case of Lady finger banana the genome formula is 1S 1C 3M 18U3 34R 287N. Similar trend was also noticed in two other banana varieties. The average BBTV genomic formula revealed that the Rep was 44 which is much higher than DNA-S, -C, and -M components and lower than DNA-U3 and -N.

Sequence analysis of Reps

Reps are known to be the most conserved amongst the different proteins of nanovirids. To examine the molecular phylogeny, comparison of Rep sequences from various members of nanovirids was undertaken along with those with small circular Rep-encoding single-stranded DNA (CRESS-DNA). A total of 138 curated sequences pertaining to four families of ssDNA viruses viz. *Nanoviridae* (14 sequences), *Alphasatellitidae* (16 sequences), *Geminiviridae* (105 sequences), and *Genomoviridae* (3 sequences) that infect plants were compiled to study the phylogenetic aspects. These sequences were procured from GenBank and from a set of 535 sequences of the Reps of ssDNA viruses belonging to the known ssDNA viral families as well as a large collection of metagenomics data [25]. The unique sequences of Rep protein were compiled, and initial multiple sequence alignment was done using Muscle [12]. The evolutionary history was inferred by using the Maximum Likelihood method and the test of phylogeny was done by the bootstrap method with 1000 bootstrap

Fig. 1 Molecular phylogeny of Rep sequences from various members of nanovirids alongside small circular Rep-encoding single-stranded DNA (CRESS-DNA). Sequence alignment of 138 curated sequences pertaining to four families of ssDNA viruses was done

replicates using MEGA6 [40]. The length of the sequences varied from 278 for Faba bean necrotic yellows C1 alphasatellite (GenBank ID: X80879) and Faba bean necrotic stunt alphasatellite (GenBank ID: KC978991) to 469 (Mungbean yellow mosaic India virus, GenBank ID: DQ400847). ESM of Table 1 shows the list of the viruses used for the analysis. Two major clades split from the root into the nanovirids group together with *Alphasatellidae* and

using Muscle [38]. The evolutionary history was inferred by using the Maximum Likelihood method and the test of phylogeny was done by the bootstrap method with 1000 bootstrap replicates using MEGA6 [34]

the Geminiviral group along with *Genomoviridae* (Fig. 1). Geminiviral sub-clade branches out into various genera viz. Mastrevirus, Capulavirus, Curtovirus, Begomovirus, Grablovirus, Eragrovirus, Turncurtovirus, and Becurtovirus based on the phylogeny of Reps. The Genomoviruses comprising of a few plant infecting members are recently given a new taxonomic position and are least represented in the present study. They clearly deviate as a separate clade

from the rest of the Geminiviruses. The nanovirids and *Alphasatellitidae* diverge into separate sub-clades, comprising of the Nanovirid members of Babuviruses and nanoviruses and the rest of the members of *Alphasatellitidae* of Geminivirids and nanovirids. *Nanoviridae* ancestry is seen to be shared by the nanoviruses, Babuviruses and Nanovirus-associated alphasatellites (subfamily *Nanoalphasatellitinae*). The genomic components of Babu- and nanoviruses that include the components DNA-R, -S, -C, -M and -N are homologous. DNAs of unidentified functions viz. DNA-U1, -U2 and -U4 that are believed to encode proteins have been identified from nanoviruses only while a similar one, DNA-U3, has been reported from babuviruses. The Reps of newer viruses emerging from the metagenomics studies such as uncultured Nanovirus (HQ335074) and Picobiliphyte sp. MS584-5 Nanovirus (GenBank ID HQ322117) show their positioning with the alphasatellites of Babu- and nanoviruses [24, 25]. The separate grouping of Babu- and Nanovirus associated alphasatellites is also quite apparent from the dendrogram (Fig. 1). The Rep of recently described Nanovirus, *Sophora* yellow stunt associated virus (SYSaV), Cow vetch latent virus (CVLV), and Black medic roll virus (BMRV) group with the Nanovirus genera of nanovirids and their alphasatellites appear to cluster with the alphasatellites of Nanovirus genera (subfamily *Nanoalphasatellitinae*). The analysis indicates a recent ancestry of nanovirids and the *Alphasatellitidae* that is distinct from that of the Gemini and *Genomoviridae* lineage. The nanovirids share very low sequence identities (17–22%) with other group II viral Rep proteins such as those of *Geminiviridae*, *Circoviridae*, and the rest though the overall domain features are conserved [51]. The positioning of Coconut foliar disease virus (CFDV) into a new taxon separate from Gemini and nanovirids also finds justice from the current analysis where the alphasatellite of the same is seen to occupy a unique place distinct from the alphasatellites of *Geminiviridae* (subfamily *Geminialphasatellitinae*) [4, 18]. As more metagenomic data become available from the plant resources for nanovirids it would be easier to put the missing links of the molecular phylogeny of Reps in place.

Alphasatellites of nanoviruses

Many satellite-like ss DNAs of about 1000–1100 nt, referred to as alphasatellites are found associated with the majority of the *Nanoviridae* isolates that encode their own respective Rep proteins [4]. They lack the CR-I and the CR-II that is characteristic of their helper viruses. While the Rep is capable of initiating replication of all the genomic segments, the Reps of alphasatellites can only initiate replication of their own respective DNA. The

alphasatellites are known to have an AT-rich domain downstream of the Rep gene and are associated with reduced infectivity [44]. The alphasatellites that are associated with *Nanoviridae* share high sequence similarity between each other and are closely related to the Reps of alphasatellites belonging to begomoviruses [4]. The nanovirus affecting *Sophora alopecuroides* in Iran comprised of four molecules of DNA-R, two molecules of DNA-C and one molecule each of DNA-S, DNA-M, DNA-N, DNA-U1, DNA151 U2 and DNA-U4 along with 14 molecules of alphasatellites [21]. The alphasatellites of SYSaV showed the consensus nonanucleotide (TAGTATTAC) found at the origins of replication in the alphasatellite DNA. Two exceptions were seen in KX534397 and KX534398 isolates that had a sequence of CAGTATTAC. Within intra-viral alphasatellites associated with SYSaV, pairwise identities ranging between 57.7 and 99.7% were observed while they were between 57.7 and 84.6% with other Nanovirus-associated alphasatellites. Phylogenetic analysis of SYSaV associated alphasatellites showed three-well supported clades accommodating all the Nanovirus-associated alphasatellites except for CFDV. The pairwise identity and phylogenetic analysis of the available alphasatellite molecules did not indicate clear association between specific groups of alphasatellites [21]. These results corroborated with the current phylogenetic analysis where the grouping of alphasatellites of nanovirids, and Geminivirids is obvious (Fig. 1). From the phylogenetic analysis it is evident that the Reps of *Alphasatellitidae* share sequence commonalities with each other and hence, are more likely to share similarities in structure and functional attributes. Consequently, the alphasatellites of nanovirids have been categorized into a separate sub-family called *Nanoalphasatellitinae* with seven different genera classified on the basis of pairwise sequence identity into *Babusatellite*, *Clostunsatellite*, *Fabenesatellite*, *Milvetsatellite*, *Mivedwarsatellite*, *Sophoyesatellite*, and *Subclovsatellite* comprising a total of 19 species derived from 54 distinct GenBank entries [4]. In all the cases, the similarity in sequence of Reps of alphasatellites with those of their respective helper viruses indicates their evolution from DNA-R during co-infection. It also suggests the necessity for the conservation of the core sequence and structural elements of the Reps that have functional significance in RCR.

The role of Rep in RCR

Following successful infection of the host cell, short endogenous DNA or RNA molecules of various sizes that are generally encapsidated together with the genomic ssDNA serve as primers for host polymerase(s) to initiate

synthesis of the complementary strand creating ds replicative form [7]. The ds intermediate serves as a template for both transcription and RCR leading to the generation of the viral sense strand that is eventually encapsidated inside the capsid to form infectious virions [39]. The viral sense strand is involved in coding for the Rep and other proteins, while the complementary strand is used as a template for rapid generation of the sense strand and hence serves as an indicator for viral replication and expression in infected cells [14, 29]. The RCR is initiated by cleavage of the phosphodiester bond in the conserved nonanucleotide sequence TATTATT¹AC or TAGTATT¹AC at the viral replication origin by Rep. In a study involving the role of three BBTV iterons in virus replication using Rep, three iterons namely the F1, F2 and R iterons were investigated. It was discovered that their roles differed from one another and any change in their sequence affected replication to a varying degree [20]. The results from in vivo experiments suggest that interaction of Rep with the F2 iteron is essential for DNA replication while the F1 and R iterons are thought to function in coordination to enhance virus replication [20]. Post nicking, the Rep becomes covalently linked to the 5'phosphate of the first adenosine through an activated tyrosine forming a phospho-tyrosyl ester that is resolved after the completion of each round of replication [41]. The catalytic tyrosine is found to be conserved in all the members of *Nanoviridae* implying its significance in initiating replication (Fig. 2). In FBNYV, infectivity of cloned viral DNAs was achieved and the origin-specific DNA cleavage and nucleotidyl transfer activities were demonstrated in vitro leading to the identification of tyrosine 79 as the catalytic residue [16, 41]. The replication machinery is assembled around the 3' end of the nick and the host DNA polymerase synthesizes a new strand after displacement of viral strand through yet unknown helicase activity using the free 3'OH at the nick as a primer. This results in the generation of concatemeric linear ss DNA producing one genome copy per turn of replication. A different Rep molecule then cleaves the newly synthesized genome copy near the inverted repeat following each round of replication. Parallely, the 3'OH of the cleaved origin sequence attacks the phospho-tyrosyl ester of the Rep linked to the 5' end of the DNA thereby releasing it and transfers it to the 3' OH in a nucleotidyl transfer reaction generating a circular ssDNA molecule [41, 42, 51] (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/ssdna-viruses-2011/w/ssdna_viruses/149/nanoviridae). In many viruses from ssDNA family there is accumulation of heterogeneous sub- and extra-genome length DNA (hDNA) during RCR that are used to generate viral RNA by the process of recombination-dependent repair by employing host ds break repair complexes.

Domain features, 3D structure, and mechanism of catalysis

The Reps have conserved domain organization with the N-terminal endonuclease and the C-terminal domain that is proposed to have helicase activity. The former is responsible for catalyzing the nicking and ligation at the origin of DNA replication during RCR, while the latter is responsible for unwinding the dsDNA intermediates. The amino-terminal endonuclease domain comprises three conserved motifs (I–III) that is common to the His-hydrophobic-His (HUH) superfamily endonucleases [23, 25] and the carboxy-terminal domain includes the conserved motifs viz. Walker A, Walker B, motif C and the 'Arginine finger' motif [9, 25]. The Rep proteins of *Nanoviridae* are characterized by the lack of retinoblastoma-like protein (Rb)-binding motif and presence of a varied dNTP-binding motif (GPQ/NGGEGKT). The Reps are believed to have a DNA binding domain endonuclease domain, a linker domain for membrane attachment and a domain for oligomerization. The amino-terminal domain of Rep comprising the first 15 residues is believed to recognize the conserved nonanucleotide motif at the apex of the stem-loop origin and carry out endonuclease activity for cleaving the DNA and initiating RCR followed by joining the DNA at the origin of replication. The carboxy-terminal domain from residue 180 onwards has ATPase activity [29]. In Geminiviruses, not only has the ATPase activity of the C-terminal domain been confirmed but also there is a central domain comprising residues 119–180 that is implicated in oligomerization of Rep [10]. These features are yet to be ascertained for the Reps of *Nanoviridae*. In the Rep protein of nanoviral FBNYV, origin-specific DNA cleavage and nucleotidyl transfer activities were demonstrated in vitro. Additionally, the involvement of divalent metal ions such as Mg²⁺ or Mn²⁺ cations in DNA cleavage was established for two different nanoviral Rep proteins [51]. The involvement of metal ions during catalysis is a common aspect of the replication of viruses with both DNA and RNA genomes [50].

In a study involving the three-dimensional solution structure of the catalytic domain of the Rep protein of FBNYV using NMR, helped gain useful insights and the in vivo characterization of its endonuclease activity [47]. The structure of the endonuclease domain of Rep from FBNYV showed a fold quite like that of HUH motif-containing relaxase proteins from conjugative plasmids that is implicated in metal binding. However, in the structure of endonuclease domain of Rep of FBNYV, there was a circular permutation with respect to the viral Rep domain, which is regular in the relaxases. The fold has the conservation of the secondary structure elements including β 1,

Fig. 2 Multiple sequence alignment of the Rep sequences of members of *Nanoviridae* procured from Uniprot. The alignment was generated with MEGA6 [40] using MUSCLE [12] and ESPRIT

(<http://espruit.ibcp.fr>) was used for displaying. The conserved catalytic tyrosine is shown with a yellow background (color figure online)

β 2, β 3, β 4, β 5, and α 2 (Fig. 3) [47]. However, certain secondary structural components such as the α 1 helix and the β 1– β 5 mini β -sheet extension were found to be absent. The endonuclease domain of FBNYV was found to be a less decorated structure although it reflected the core structure for all viral Rep endonuclease domains known to

date [22]. The solution structure of the endonuclease domain of FBNYV was used to propose a model for the basic mechanism of catalysis [47]. It was proposed that the binding and positioning of the phosphate group occurred at the scissile bond with the aid of the divalent metal ion and the basic side chain of the conserved lysine leading to the

Fig. 3 The fold of the endonuclease domain of FBNYV (PDB id: 2HWT) colored according to the secondary structure and shown in ribbon representation. The figure was generated using Chimera (Pettersen et al. [35])

polarization of the phosphodiester bond. This, in turn, led to the formation of a penta-coordinated phosphate intermediate due to the nucleophilic attack of the hydroxyl group of the catalytic tyrosine residue on the phosphate. Finally, the proton of the hydroxyl group from the catalytic tyrosine was abstracted by histidine residues coordinated by metal ions leading to the formation of the phosphotyrosine bond and a covalent adduct between the protein and the 3'-fragment of the substrate DNA [47]. Prior to the cleavage reaction, the conserved nonanucleotide in the origin are recognized by a cluster of positively charged residues of the Rep endonuclease domains as demonstrated in studies involving Tomato yellow leaf curl Sardinia virus (TYLCSV, [6] Porcine circovirus type 2 (PCV2), [48] and Adeno associated virus 5 (AAV5) [22]. EMSA studies on complexes of PCV2 is suggestive of similar modes of DNA recognition and binding for Rep proteins of *Nanoviridae* during RCR involving close contacts with the exposed side of the central sheet close to the catalytic residues [47]. A crucial combination of high-affinity binding of ds DNA sequences at the origin with Rep coupled with a transient and low-affinity metal-mediated binding of the ss region is proposed to repel the 5' part of the nonanucleotide sequence following cleavage leading to the displacement of this sequence from the protein that generates the primer

necessary for the further synthesis of DNA during RCR [47].

Structural comparison of Reps

The three-dimensional superposition of the structures of known Reps from ssDNA viruses available at PDB (www.rcsb.org) [1] was done using Multiprot server [37]. Four available structures were used for the study, two from the PCV2 and one each from TYLCSV and FBNYV. Table 1 summarizes the results of superposition. The alignment size was 54 and the root mean square deviation of the structures was 1.38 Å. Nanoviral Reps are found to be smaller in size to their counterparts in Gemini- and Circoviruses. Overall, the structural similarity of the endonuclease domain of individual Reps is striking between the otherwise diverse families of ssDNA viruses including nanovirids [47], Geminivirids [6] and Circovirids [48]. The core β strands are similar in length and disposition to the Nanovirid Reps, all of them comprising of the central four-stranded antiparallel β sheets and the flanking two helices. While the regions pertaining to α A and α C are replaced by long loops of 14 and 23 residues, respectively, in nanovirids, the α B is lacking in Geminiviruses and comprises of a loop of 26 residues connecting β 3 and β 4 but traces the same path as the α B of nanovirids and Circoviruses (Fig. 4a). The amino and carboxy-terminal tails of the domains, the mini β -sheet extensions of the Gemini- and Circovirus Rep proteins, and the corresponding region of the nanoviral Rep are seen to contribute to the positively charged area involved in recognition of the origin. The b-factors of the C α atoms were plotted for the PCV2 and FBNYV and they follow a uniform trend for both the structures (Fig. 4b). It is noted that there is an overall high thermal factor of the atoms especially around the loop and the terminal regions of the structure. Among the conserved motifs of the endonuclease domain, the motif I comprises the residues Cys–Phe–Thr–Leu in FBNYV that are in the

Table 1 Comparison of the three-dimensional structures of the endonuclease domains of Rep from ssDNA viruses

Virus (<i>family</i>)	PDB ID	No. of C- α	No. of residues									Motifs		
			β 1	β 2	β 3	β 4	β 5	α A	α B	α C	I	II	III	
Tomato yellow leaf curl virus-Sardinia (<i>Geminiviridae</i>)	112 m	118	6	10	9	5	3	10	–	8	FLTY	HUH	YXXK	
Porcine circovirus type 2 (<i>Circoviridae</i>)	2hw0	115	8	12	8	4	6	8	8	11	VFTL	HUQ	YXXK	
Porcine circovirus type 2 (<i>Circoviridae</i>)	5xor	98	8	12	8	4	6	8	8	11	VFTL	HUQ	YXXK	
Faba bean necrotic yellow virus (<i>Nanoviridae</i>)	2hwt	94	6	8	8	3	5	–	7	–	CFTL	HUQ	YXXKE/ D	

The PDB structures were downloaded from www.rcsb.org. Except for TPDB ID 5xor that is a crystal structure, all the others are solution structures solved using NMR

Fig. 4 **a** Structural superposition of the endonuclease domain of individual Reps between the otherwise diverse families of ssDNA viruses including nanovirids (Faba Bean Necrotic Yellow Virus, FBNYV) shown in orange, Geminivirids (Tomato yellow leaf curl Sardinia virus, TYLCSV) shown in green, and Circovirids (Porcine

circovirus type 2, PCV2) shown in blue. **b** The Superposition showing the conservation of catalytic tyrosine in the different Rep structures from the three families. The catalytic residues of Motif III are shown as sticks (color figure online)

β 1 strand where the Cys and the Thr face the exposed surface of the protein while Phe and Leu face the core hydrophobic region. The equivalent residues in TYLCSV are Phe–Leu–Thr–Tyr and in PCV2 are Val–Phe–Thr–Leu. It has been earlier suggested that the hydrophobic residues of the motif I (the first and the third residues) contribute to the stability and architecture of the domain while the exposed residues (the second and the fourth) aid in the recognition of nucleic acid. The motif II is located in the central β 3 strand and comprises of HUH for Geminivirus, HUQ for Nano and Circovirus. The residue U in the motif represents a large, hydrophobic amino acid that is buried in the hydrophobic core while the flanking H are Histidines and Q represents Glutamine that are implicated in divalent metal coordination. The motif III, YXXKE/D, is located at the catalytic helix and contains the active site tyrosine. The Tyr and Lys of the motif point towards the exposed side of the β -sheet in all the viruses and indicate a similar catalytic scheme for all the Reps (Fig. 4c). Overall the Reps of nanovirids show greater structural similarities to the Reps of Circoviruses when compared to Geminiviruses. This supports the previous proposal that Circoviruses might have evolved via recombination events between a picorna-like virus with ss RNA genome and a Nanovirus [15, 32].

Interactions of Rep

Earlier studies with viruses such as Wheat dwarf virus (WDV) [30], TYLCSV [8], and Tomato golden mosaic virus (TGMV) [26, 27] helped in identifying crucial

interaction partners of the Rep proteins involved in replication. These included the Replication factor C protein, retinoblastoma-related protein, sumoylation enzyme, NbSCE1 etc. Many proteins that are seen to interact with proteins of group II viruses of *Nanoviridae* were identified using the yeast two-hybrid system [46]. In vitro pull-down or immunoprecipitation techniques were further employed to confirm the interacting partners. *In planta* expressed and affinity purified oligohistidine-tagged Rep variants of FBNYV were proficient in catalyzing initiation of viral DNA replication and termination in *Nicotiana benthamiana*. Further, it was observed that the artificial replicon encoding the tagged Rep protein could spread throughout the plant and engage in aphid-mediated transmission when other essential proteins such as the movement protein, capsid protein etc. were provided by coinfection with a helper virus [46]. The oligohistidine-tagged Rep encoding DNAs were demonstrated to be tightly associated with virus capsid protein through IC-PCR studies and were also proposed to be encapsidated into true virions. It had been reported earlier that geminivirus Rep proteins formed oligomers in solution in vitro in a pH-dependent fashion [19, 33]. The rep protein of nanoviruses was proposed to similarly oligomerize but is yet to be fully confirmed. Studies with his-tag proteins of Rep of FBNYV showed interactions with rep indicating the possibilities of oligomerization [46]. In a recent study cellular localization and inter-protein interactions of PNVDV were analyzed by GFP tagging and bimolecular fluorescence complementation, yeast two-hybrid or in vitro analyses, which demonstrated the interactions of Reps with each other suggesting

the possibility of oligomerization. The Rep was seen to be uniformly distributed in the nucleus and is predicted to carry a nuclear localization signal (NLS), “RKRR”-motif, at the central region. The pairwise interactions of Rep with other proteins also suggested the possibility of strong interactions with CP and NSP [28]. While the studies involving interactions of Geminiviral Reps with self and host proteins have been actively undertaken, such studies are lacking with respect to Nanovirid Reps and mandate immediate attention.

Future considerations

The Nanovirid Reps show a greater sequence divergence over structural aspects. Availability of limited structural information has handicapped the elucidation of molecular details of RCR meted out by the Reps. Further, the knowledge of the structural aspects of helicase action and oligomerization is also lacking. Details of proteomics-based interaction studies of Reps with self and host proteins would also go a long way in understanding the molecular mechanism of initiation of RCR, capsid assembly and systemic spread.

References

- Berman H, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucl Acids Res.* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235>.
- Biagini P. Classification of TTV and related viruses (Anelloviruses). In: de Villiers E-M, Zur H, editors. *TT viruses: the still elusive human pathogens*. Berlin: Springer; 2009. p. 21–33.
- Breitbart M, Delwart E, Rosario K, Segales J, Varsani A, Consortium IR. ICTV virus taxonomy profile: Circoviridae. *J Gen Virol.* 2017;98:1997–8. <https://doi.org/10.1099/jgv.0.000871>.
- Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Elvira F-O, Moriones E, Lett J-M, Zerbini FM, Varsani A. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus—and nanovirus—associated alphasatellites. *Arch Virol.* 2018. <https://doi.org/10.1007/s00705-018-3854-2>.
- Burns TM, Harding RM, Dale JL. The genome organization of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol.* 1995;76:1471–82.
- Campos-Olivas R, Louis J, Clerot D, Gronenborn B, Gronenborn A. The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc Natl Acad Sci USA.* 2002;99:10310–5.
- Can MC, Rodri EA, Moriones E, Bejarano ER, Grande-Pe A. A sensitive method for the quantification of virion-sense and complementary-sense DNA strands of circular single-stranded DNA viruses. *Sci Rep.* 2014;4:6438. <https://doi.org/10.1038/srep06438>.
- Castillo AG, Collinet D, Deret S, Kashoggi A, Bejarano ER. Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology.* 2003;312:381–94. [https://doi.org/10.1016/S0042-6822\(03\)00234-4](https://doi.org/10.1016/S0042-6822(03)00234-4).
- Choi I, Stenger D. Strain-specific determinants of beet curly top geminivirus DNA replication. *Virology.* 1995;206:904–12.
- Clerot D, Bernardi F. DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. *J Virol.* 2006;80:11322–30. <https://doi.org/10.1128/JVI.00924-06>.
- Cotmore SF, Davison AJ. The family Parvoviridae. *Arch Virol.* 2015;159:1239–47. <https://doi.org/10.1007/s00705-013-1914-1>.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res.* 2004;32:1792–7. <https://doi.org/10.1093/nar/gkh340>.
- Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge J, Chang H, Doszt Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings D, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Xenarios I, Yeh L, Young S, Mitchell AL. InterPro in 2017—beyond protein family and domain annotations. *Nucl Acids Res.* 2017;45:190–9. <https://doi.org/10.1093/nar/gkw1107>.
- Fondong V. Geminivirus protein structure and function. *Mol Plant Pathol.* 2013;14:635–49.
- Gibbs MJ, Weiller GF. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc Natl Acad Sci.* 1999;96:8022–7.
- Grigoras I, Timchenko T, Gronenborn B. Transcripts encoding the nanovirus master replication initiator proteins are terminally redundant. *J Gen Virol.* 2008. <https://doi.org/10.1099/vir.0.83352-0>.
- Grigoras I, Timchenko T, Gronenborn B. Transcripts encoding the nanovirus master replication initiator proteins are terminally redundant. *J Gen Virol.* 2008;89:583–93. <https://doi.org/10.1099/vir.0.83352-0>.
- Gronenborn B, Randles JW, Knierim D, Barrière Q, Vetten HJ, Warthmann N, Cornu D, Sileye T, Winter S, Timchenko T. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. *Sci Rep.* 2018. <https://doi.org/10.1038/s41598-018-23739-y>.
- Gutierrez C, Ramirez-Parra E, Castellano MM, Sanz-Burgos AP, Luque A, Missich R. Geminivirus DNA replication and cell cycle interactions. *Vet Microbiol.* 2004;98:111–9. <https://doi.org/10.1016/j.vetmic.2003.10.012>.
- Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. *J Gen Virol.* 2006;87:3409–12. <https://doi.org/10.1099/vir.0.82166-0>.
- Heydarnejad J, Kamali M, Massumi H, Kvarnheden A, Male MF, Kraberger S, Stainton D, Martin DP, Varsani A. Identification of a nanovirus–alphasatellite complex in *Sophora alopecuroides*. *Virus Res.* 2017;235:24–32. <https://doi.org/10.1016/j.virusres.2017.03.023>.
- Hickman AB, Ronning DR, Perez ZN, Kotin RM, Dyda F. The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol Cell.* 2004;13:403–14.
- Horser C, Harding R, Dale J. Banana bunchy top nanovirus DNA-1 encodes the “master” replication initiation protein. *J Gen Virol.* 2001;82:459–64.
- Kazlauskas D, Dayaram A, Kraberger S, Goldstien S. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses.

- Virology. 2017;504:114–21. <https://doi.org/10.1016/j.virol.2017.02.001>.
25. Kazlauskas D, Varsani A, Krupovic M. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. *Viruses*. 2018;10:1–11. <https://doi.org/10.3390/v10040187>.
 26. Kong L, Hanley-Bowdoin L. A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. *Plant Cell*. 2002;14:1817–32. <https://doi.org/10.1105/tpc.003681.AL1>.
 27. Kong L, Orozco BM, Roe JL, Nagar S, Ou S, Feiler HS, Durfee T, Miller AB, Gruissem W, Robertson D, Hanley-Bowdoin L. A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO J*. 2000;19:3485–95.
 28. Krenz B, Schieff I, Greiner E, Krapp S. Analyses of pea necrotic yellow dwarf virus-encoded proteins. *Virus Genes*. 2017. <https://doi.org/10.1007/s11262-017-1439-x>.
 29. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucl Acids Res*. 2018;46:708–17. <https://doi.org/10.1093/nar/gkx932>.
 30. Luque A, Sanz-Burgos P, Ramirez-Parra E, Castellano MM, Gutierrez C. Interaction of geminivirus Rep protein with replication factor C and its potential role during geminivirus DNA replication. *Virology*. 2002;302:83–94. <https://doi.org/10.1006/viro.2002.1599>.
 31. Mandal B. Advances in small isometric multicomponent ssDNA viruses infecting plants. *Indian J Virol*. 2010;21:18–30. <https://doi.org/10.1007/s13337-010-0010-3>.
 32. Martin DP, Biagini P, Lefevre P, Golden M, Roumagnac P, Varsani A. Recombination in eukaryotic single stranded DNA viruses. *Viruses*. 2011;3:1699–738. <https://doi.org/10.3390/v3091699>.
 33. Missich R, Ramirez-Parra E, Gutierrez C. Relationship of oligomerization to DNA binding of wheat dwarf virus RepA and Rep proteins. *Virology*. 2000;188:178–88. <https://doi.org/10.1006/viro.2000.0412>.
 34. Mware BO. Development of Banana bunchy top virus resistance in bananas: RNAi approach. Ph.D. thesis. Queensland University of Technology; 2016.
 35. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualisation system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605–12.
 36. Selvarajan R, Mandal B, Balasubramanian V, Banerjee A, Vijayanandraj S, Ghosh A. Biology and molecular biology of babuviruses occurring in India. In: Mandal B, Rao GP, Baranwal V, Jain R, editors. A century of plant virology in India. Singapore: Springer; 2017. pp. 27–48.
 37. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. *Proteins*. 2004;56:143–56. <https://doi.org/10.1002/prot.10628>.
 38. Sicard A, Michalakakis Y, Gutiérrez S, Blanc S. The strange life-style of multipartite viruses. *PLoS Pathog*. 2016;12:1–19. <https://doi.org/10.1371/journal.ppat.1005819>.
 39. Stenlund A. Initiation of DNA replication: lessons from viral initiator proteins. *Nat Rev Mol Cell Biol*. 2003;4:777–85.
 40. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9. <https://doi.org/10.1093/molbev/mst197>.
 41. Timchenko T, de Kouchkovsky F, Katul L, David C, Vetten HJ, Gronenborn B. A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *J Virol*. 1999;73:10173–82.
 42. Timchenko T, Katul L, Sano Y, De Kouchkovsky F, Vetten HJ, Gronenborn B. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. *Virology*. 2000;274:189–95. <https://doi.org/10.1006/viro.2000.0439>.
 43. Timchenko T, Katul L, Sano Y. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. *Virology*. 2000;195:189–95. <https://doi.org/10.1006/viro.2000.0439>.
 44. Timchenko T, Katul L, Aronson M. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. *J Gen Virol*. 2006;87:1735–43. <https://doi.org/10.1099/vir.0.81753-0>.
 45. Varsani A, Krupovic M, Krupovic M. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Arch Virol*. 2018. <https://doi.org/10.1007/s00705-018-3820-z>.
 46. Vega-Arregui JC, Timchenko T, Gronenborn B, Ramirez BC. A functional histidine-tagged replication initiator protein: implications for the study of single-stranded DNA virus replication in plants. *J Virol*. 2005;79:8422–30. <https://doi.org/10.1128/JVI.79.13.8422>.
 47. Vega-Rocha S, Gronenborn B, Gronenborn AM, Campos-Olivas R. Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus Faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circovirus structures. *Biochemistry*. 2007;46:6201–12.
 48. Vega-Rocha S, Byeon IJL, Gronenborn B, Gronenborn AM, Campos-Olivas R. Solution structure, divalent metal and DNA binding of the endonuclease domain from the replication initiation protein from porcine circovirus 2. *J Mol Biol*. 2007;367:473–87. <https://doi.org/10.1016/j.jmb.2007.01.002>.
 49. Venkataraman S, Savithri HS, Murthy MRN. Structural aspects of plant viruses. In: Gaur RK, Khurana SMP, Dorokhov Y, editors. Plant viruses diversity, interaction and management; CRC Press, Taylor and Francis Group 2018. pp. 3–26.
 50. Venkataraman S, Prasad BVLS, Selvarajan R. RNA dependent RNA polymerases: insights from structure, function and evolution. *Viruses*. 2018;10:76. <https://doi.org/10.3390/v10020076>.
 51. Vetten H. Nanoviruses. In: Mahy BW, Van Regenmortel MHV, editors. Encyclopedia of virology. 3rd ed. Oxford: Elsevier; 2008. p. 385–91.
 52. Virol A, Krupovic M. Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch Virol*. 2016;1:8. <https://doi.org/10.1007/s00705-016-2943-3>.
 53. Wanitchakorn R, Harding R, Dale J. Banana bunchy top virus DNA-3 encodes the viral coat protein. *Arch Virol*. 1997;142:1673–80.
 54. Zerbini FM, Bridson RW, Idris A, Martin DP, Moriones E, Navas-Castillo J, Rivera-Bustamante R, Roumagnac P, Varsani A, Consortium IR. ICTV ICTV virus taxonomy profile: Geminiviridae. *J Gen Virol*. 2017;98:131–3. <https://doi.org/10.1099/jgv.0.000738>.